v-LEARNING
Using Voice for Distant Learning in Emerging Regions

Thushari Atapattu and Kasun De Zoysa
University of Colombo, School of Computing, 35, Reid Avenue, Colombo 7, Sri Lanka

Keywords: WWTW, VoiceXML, Transcoder, Asterisk.

Abstract: At present, accessing the internet through visual interfaces is the most common approach. However, it requires some basic resources such as a computer or web-enabled mobile device, an internet connection, electricity and some amount of IT literacy. Because of the relatively high cost of this set up, underprivileged users are unaware or have no direct access to the internet. Since voice communications through telephony systems do belong to the growing trend, people make use of telephones for various purposes. Among them, accessing web through telephone devices is explored. In this paper, we present an approach to access the learning materials of the Learning Management System (LMS) of University of Colombo School of Computing, Sri Lanka through interactive voice driven applications.

1 INTRODUCTION

World Wide Web (WWW) has become the major information source around the world. People access the web for various purposes such as learning, communication, entertainment etc. As a result, web has grown to be one of the most popular media in the world. But, in order to access the internet, it requires essential resources. Basically, it needs a computer (or modern mobile device, internet kiosks etc.), an internet connection (broadband, wire-less, dial-up etc.), a telephone, electricity and some fundamental IT skills.

Buying a computer and obtaining a fixed internet connection might cost around LKR 80,000 (US$800), which is unaffordable for many people in developing countries. At the same time, such internet connections are not available in the rural areas. Majority of the people in developing countries do not even have electricity in their homes. Due to lack of resources, it is an overhead for a majority of people in the world to access internet directly.

According to the internet world stat (Internet world stat, 2008), approximately 22% of the people in the world have access to the internet (Table 1). The table 1 shows that most of the emerging regions such as Africa and Asia have limited access to the web. It implies that most of the people in the world (78%) are still untouched to these sophisticated facilities.

Table 1: Internet Usage and world population statistics for June 30, 2008.

World Region	Population	% Population (Penetration)	% Usage Growth 2000-2008
Africa	955,206,348	5.3%	1031.2%
Asia	3,776,181,949	15.3%	406.1%
Europe	800,401,065	48.1%	266%
North America	337,167,248	73.6%	129.6%
Middle East	197,090,443	21.3%	1176.8%
Latin America/Caribbean	576,091,673	24.1%	669.3%
Oceania/Australia	33,981,562	59.5%	165.1%
World Total	6,676,120,288	21.9%	305.5%

According to the statistics, internet penetration is around 2.2% in Sri Lanka, 2007 (Sri Lanka Internet world stat, 2008). It is relatively a low rate with compared to other educational facilities in Sri Lanka. The situation is raised due several reasons. The most critical issue is the unavailability of electricity and internet connectivity facilities in rural areas. In order to improve these factors, an infrastructural development should be done which requires a huge investment.

Atapattu T. and De Zoysa K. (2009).
v-LEARNING - Using Voice for Distant Learning in Emerging Regions.
In *Proceedings of the First International Conference on Computer Supported Education*, pages 216-221
DOI: 10.5220/0001971702160221

As an alternative, government and private sectors have invested on public internet accessible places such as Internet cafes, Public Internet Kiosk etc. Some of these solutions are still not possible for underprivileged users as these solutions are also costly.

Table 2: Internet usage and population statistics in Sri Lanka.

Year	User	Population	%penetration
2000	121,500	19,630,230	0.5%
2007	428,000	19,796,874	2.2%

Since mobile technologies have been rapidly growing, people make use of their mobile devices to access the web anywhere in the world. This will reduce the overhead of buying a PC with an internet connection and the expenditure for electricity. Accessing wire-less networks also requires high charges, which is not affordable for the people in developing regions.

Apart from that, accessing internet through mobile devices also has some additional limitations. Firstly, buying a featured phone is not less expensive in Sri Lanka. At the same time, the GSM/GPRS coverage does not exist in rural areas. The mobile device is not always user friendly since it has a small screen and a tiny keypad. All of these limitations will distract people accessing the web through their hand-held devices.

All the above mentioned techniques require browsing the internet through a visual interface such as a web browser. Due to the above mentioned limitations, some researchers have explored the possibility of accessing the web through voice communication. The basic voice communication has had a larger penetration among the world population as well as in Sri Lanka. Therefore, IBM Research Laboratory (Kumar, 2007) has conducted a research, which uses voice to access the internet. This concept is called World Wide Telecom Web (WWTW) (Kumar, 2007). In this model, the voice sites are developed instead of typical web sites. Those voice sites are implemented using a language called VoiceXML (VoiceXML, 2008). VoiceXML is a markup language derived from XML. Users are allowed to call to the voice site which is a collection of VoiceXML pages.

The preliminary attempt of this work is to build an interactive voice learning environment for the undergraduates of University of Colombo School of Computing (UCSC). Since the cost of basic voice communication through telephone is relatively low, accessing web using voice is encouraged. This will be beneficial for underprivileged students who have no direct access to the teaching and learning materials in the web.

This paper is organized as follows. In section 2, the work related to World Wide Telecom Web is discussed. Our proposed architecture and overview of the system is detailed in section 3. The system functionalities are explained in section 4. Finally the proposed system is summarized in section 5.

2 RELATED WORK

WWTW (Kumar, 2007) is a tremendous concept of IBM India Research Laboratory, where voice-driven eco systems are developed parallel to that of the WWW. The approach enables deprived population to become a part of the networked world through low cost voice communication. This concept was the basement for various researches related to voice-enabled applications.

Interactive Voice Response (IVR) systems are currently most widely used voice-driven applications in the world. Air-line, hotel reservations, telecom service providers commonly use these fixed menu-driven, user input (DTMF) based applications. These automated systems require high investments and it is not supportable for non-profit organizations and the government education sector.

Researchers have developed a low cost IVR by integrating the existing open source applications and tools (King, 2006). This system is a hybrid of OpenVXI (Carter, 2002) and Asterisk (Asterisk, 2008). OpenVXI (Carter, 2002) is a VoiceXML interpreter developed by speech group at CMU. It provides APIs for speech synthesis, speech recognition and telephony services. Asterisk (Asterisk, 2008) is the mostly used opensource PBX system in non-commercial applications and Voiceone (VoiceOne, 2008) is the web based GUI for Asterisk PBX. The gateway can be utilized to replace the existing high cost IVR systems.

VOIGEN (Kumar, 2007) enables telephone subscribers to access voice-driven systems through ordinary telephone lines. It permits individuals to create, host and deploy customized voice driven services. VOISERV (Kumar, 2007) is similar to VOIGEN (Kumar, 2007) where VOIGEN (Kumar, 2007) create and deliver data services and VOISERV (Kumar, 2007) delivers converged services. Both the systems create their own customized voicesites.

The IBM WebSphere Transcoding Publisher (WTP) (Lamb, 2008) is a commercially available product that can be used to convert HTML to VoiceXML. A group from Virginia Tech, VA has conducted a research to transcode HTML to VoiceXML using annotations (Shao, 2003).

3 PROPOSED SYSTEM

In order to experiment voice based solutions for distant learning, we have proposed an interactive voice driven system which is explained in this section. Our proposed approach will be developed in 3 stages as listed below,

1. A voice component which gives access to practice quizes in the Learning Management System (LMS)
2. Voice site parallel to existing UCSC Learning Management System (http://www.ucsc.cmb.ac.lk/lms)
3. Voice module for open source moodle project (http://www.moodle.org)

At present, we are in the process of developing the first stage of the system. In order to provide voice based access to practice quizzes in the LMS, we have implemented a simple automated Moodle XML (MoodleXML, 2008) to the VoiceXML (VoiceXML, 2008) converter. Moodle XML (MoodleXML, 2008) is a XML based language which follows XML standards. The quizes of the Learning Management System could be exported as Moodle XML. Our converter simply converts the Moodle XML files to VoiceXML files. The converted VoiceXML files are intended to interpret through VoiceXML Interpreter (Carter, 2002).

In the second stage of our proposed project it is expected to build a voice site in parallel to the existing UCSC LMS. This would be fully automated system generated from web system. The voice site is intended to be updated automatically with respect to the web system. At the final stage of the project, we have proposed to build a voice module for the open source moodle project. This would be beneficial to the society, as the people are used to customize moodle for their learning and teaching purposes.

The main focus of our proposed system is to allow voice access to learning materials for the UCSC undergraduate, external and postgraduate students. The system can be sub divided into three main components.

1. Private Branch Exchange (Asterisk server and soft phone)
2. VoiceXML Transcoder
3. VoiceXML Interpreter

The Figure 1 depicts the overall architecture of the system and each of the above sub components will be discussed in sub sections.

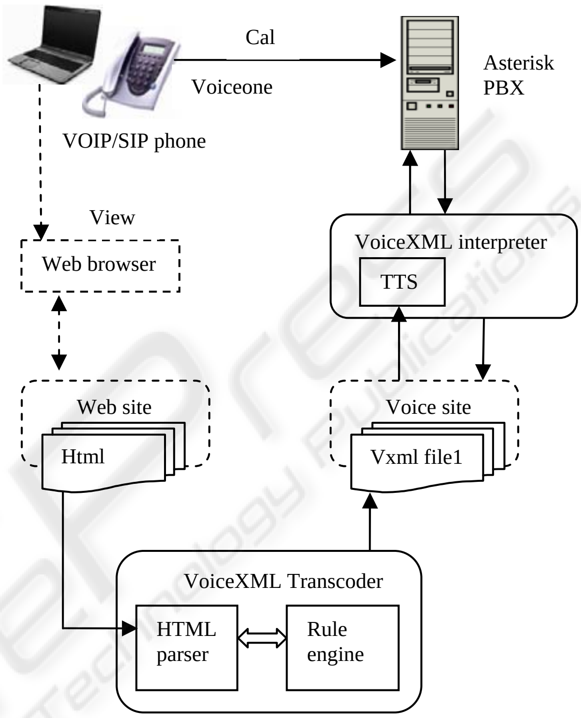

Figure 1: Overall architecture of the system.

3.1 Private Branch Exchange (PBX)

Private Branch Exchange is a telephone exchange, which serves a particular set of people. It could be located in a company, school, university etc. The cost of deploying a commercial PBX system is very high. Accordingly, we have used an open source PBX engine called Asterisk server (Asterisk, 2008) for our project.

Besides, one of the latest trends in PBX development is the Voice Over IP (VOIP) PBX, where internet protocols are used to communicate. The initial focus of the development is to configure a SIP phone to connect with the Asterisk server. For this purpose, we have used the freely available Ekiga (Ekiga, 2008) soft phone. A typical PBX set up is shown in Figure 2 below.

The asterisk server is basically capable to,

1. Get the user’s input
2. Interactively provide voice response
3. Call forwarding to voice sites

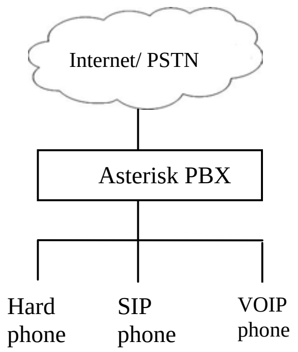

Figure 2: Overview of Private Branch Exchange.

3.2 VoiceXML Transcoder

It is a known fact that web pages are implemented using HTML. Likewise, voice pages have been built using a language called VoiceXML (VoiceXML, 2008). As the HTML pages are interpreted visually through web browsers, VoiceXML files are interpreted using voice browsers. For that the system should generate voice pages or convert existing web pages to voice pages.

The main objective of the system is to implement a voice site in parallel to the existing UCSC LMS web site. In order to do that, selected HTML web pages from the LMS site should be converted to voice pages. This process could be done through “HTML to VoiceXML Transcoder”. As there are no any open source VoiceXML transcoders available, our system is expects to implement a VoiceXML transcoder from the scratch. Our proposed transcoder has 3 main components.

1. HTML parser
2. VoiceXML translator
3. Rule engine

The overview of the proposed transcoder is shown in Figure 3.

Firstly, the static HTML pages are analyzed through a HTML parser and a HTML node tree will be generated. Once the structure of HTML node tree is analyzed, the page is converted into a VoiceXML file internally by the system.

When applying the transcoding logic, our system makes use of grammar rules which have been defined by us. After validating the conversion with the rule engine, the syntactically correct VoiceXML file will be created.

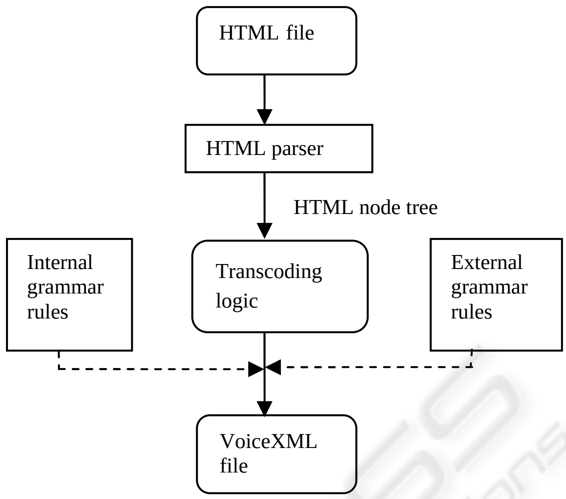

Figure 3: Overview of HTML to VoiceXML Transcoder.

A simple HTML file and it’s corresponding VoiceXML file is shown in Figure 4 below.

```
<html>
<head>
<title>Welcome to University of Colombo School of
Computing</title>
</head>
<body>
<form next=”method”
<INPUT type=”radio” name=”degree” value=”computer
science”>computer science<br>
<INPUT type=”radio” name=”degree”
value=”ICT”>ICT<br>
</form>
</body>
</html>
```

```
<?xml version=”1.0”>
<vxml version=”2.0”>
<form>
<prompt>Welcome to University of Colombo School of
Computing</prompt>
<field name=”degree”>
<prompt>Select your degree<enumerate/>
</prompt>
<option dtmf=”1” VALUE=”computer science”>computer
science</option>
<option dtmf=”2” VALUE=”ICT”>ICT</option>
</field>
</form>
</vxml>
```

Figure 4: Simple HTML and VoiceXML file.

3.3 VoiceXML Interpreter

XML based languages require an interpreter to interpret the markup commands. Accordingly, VoiceXML files should be interpreted automatically after the file is altered. OpenVXI (Carter, 2002) is one of the freely available VoiceXML interpreter used by majority of voice application builders.

Typical VoiceXML interpreter consists of 3 sub components.

1. Text-to-speech system (TTS)
2. Voice recognition system
3. User action handler

3.3.1 Text-to-Speech System (TTS)

Text-to-speech system is a way to present text output to the user through voice communication. In our system, we are using an open source TTS called FreeTTS (FreeTTS, 2008). It extracts the output from the VoiceXML file and presents it to the user through a soft phone.

3.3.2 Speech Recognition System

A typical voice-driven application has a component to recognize user's speaking context. In our proposed system, we have omitted this component and instead we are collecting user's input through Dual-Tone-Multi-Frequency (DTMF). The system prompts choices for the user and based on these choices, user has to select a number which can be entered through a telephone dial pad.

3.3.3 User Action Handler

This component is capable of collecting user's input and respond accordingly. For instance, if the user does not perform any action at his turn, the interpreter gives him a second chance to try the commands or inform him to end the call. Moreover, user action handler collects user inputs which are given by the dial pad. Likewise user action handler automatically performs several intermediate actions like a human being.

4 SYSTEM FUNCTIONALITIES

In this section, we describe the main functionalities of the system. The proposed approach is intended to be accessible via interactive telephone communication only. The user should make a call to the system in order to access the contents. The system is automated to provide services to the user regardless of other matters. The functionalities of the system can be categorized into 2 subsections as follows,

1. User-level functionalities
2. System-level functionalities

4.1 User-level Functionalities

In order to get the benefits from the distant learning project, the user should place a call to the system. This can be done through the dedicated telephone number which is assigned to the voice site. The user’s call would then automatically be handled by the Asterisk (Asterisk, 2008) server, where voiceone (VoiceOne, 2008) is the front end of the server.

The system identifies the call and redirects it to an appropriate voice site (At present, we have only one voice site in our system). The system prompts information to the user and gets their inputs through DTMF.

A sample user-system interaction is given below,

- User places a call to the system through the voice number given.
- System: Welcome to Learning Management System of University of Colombo School of Computing. Main Menu, For site news press 1, For undergraduate courses press 2, For Examinations press 3, For inquiries press 4, To exit from the system just Hang-up etc.
- User enters 3 through DTMF
- System: You have selected examinations. For Time table press 1, For exam results press 2. To go to the main menu press 0 etc.

The user can navigate through sub menus for his destination or simply can exit from any menu or sub menus. If the user fails to respond to the system within a given time frame, the menu (or sub menu) will be repeated once. If the user does not respond to the system further, the conversation will be disconnected automatically.

4.2 System-level Functionalities

At the system level, VoiceXML files will be generated and updated dynamically. This could be done by converting existing HTML files. The collection of VoiceXML files is integrated as Voicesites. The VoiceXML interpreter then interprets these VoiceXML files and presents them to the user through the TTS according to their requests. Before the voice prompts are presented to the user, VoiceXML files will be validated through the system.

5 CONCLUSIONS

The V-learning project is proposed for the underprivileged users to provide access for learning resources through the voice communication. In our approach, we have explored the concept of World Wide Telecom Web that would be parallel to that of the World Wide Web. The motivation of our approach is to deliver the services for the benefit of the students in developing economies. Though it has several benefits such as low cost, it would not be as attractive as graphical user interfaces. We believe that the system would be a bridge between the IT-savvy and the non-IT-savvy population in the world.

ACKNOWLEDGEMENTS

The author is grateful to the UCSC e-learning centre for funding the V-Learning project and the WASN laboratory for providing resources to conduct the research.

REFERENCES

King, A., Terzoli, A., Clayton, P., 2006. Creating a low cost VoiceXML Gateway to replace IVR systems for rapid deployment of voice applications. *In proceedings of the Southern Africa Telecommunication Networks and Applications conference.*

Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S. K., Nanavati, A. A., 2007. WWTW: The World Wide Telecom Web. *In Proceedings of the 2007 workshop on Network systems for developing regions, Kyoto, Japan.*

Carter, J., Eberman, B., Goddeau, D., Meyer, D. 2002. Building VoiceXML Browsers with OpenVXI. *In proceedings the international WWW conference.*

Kumar, A., Rajput, N., Chakraborty, D., Jindal, S., Nanavati, A. A., 2007. VOIGEN: A Technology for Enabling Data Services in Developing Regions. *IBM Research Report No.RI0700.*

Shao, Z., Capra, R., Perez-Quinones, M. A., 2003. Annotations for HTML to VoiceXML Transcoding: Producing Voice WebPages with Usability in Mind.

Kumar, A., Rajput, N., Chakraborty, D., Jindal, S., Nanavati, A. A., 2007. VOISERV: Creation and Delivery of Converged Services through Voice for Emerging Economies.

World Internet usage Statistics News and World Population stats, viewed 17 September 2008, <http://www.internetworldstats.com/stats.htm>

Voice Extensible Markup Language (VoiceXML) Version 2.0, viewed 3 September 2008 <http://www.w3.org/TR/voicexml20/>

Sri Lanka Internet Usage and Telecommunication reports, viewed 8 October 2008 <http://www.internetworldstats.com/asia/lk.htm>

Asterisk: The Open Source PBX & Telephony platform, viewed 13 November 2008 < http://www.asterisk.org/>

Ekiga.net :What are the service provided by Ekiga.net?, viewed 26 November 2008, <http://www.ekiga.net>

Software Voip VoiceOne- Open source PBX on Asterisk.- PBX software Voip, viewed 3 December 2008, <http://www.voiceone.it/>

Lamb, M., Horowitz, B. Guidelines for a VoiceXML Solution Using WebSphere Transcoding Publisher, viewed 17 November 2008, <http://www-3.ibm.com/software/webservers/transcoding/library.html>

FreeTTS 1.2- A speech synthesizer written entirely in the java(TM) programming language, viewed 22 November 2008 < http://freetts.sourceforge.net/docs/index.php>

Moodle XML format- MoodleDocs, viewed 19 October 2008, <http://docs.moodle.org/en/Moodle_XML>